

# Generalized Assorted Pixel Camera: Postcapture Control of Resolution, Dynamic Range, and Spectrum

Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K. Nayar, *Member, IEEE*

**Abstract**—We propose the concept of a generalized assorted pixel (GAP) camera, which enables the user to capture a single image of a scene and, after the fact, control the tradeoff between spatial resolution, dynamic range and spectral detail. The GAP camera uses a complex array (or mosaic) of color filters. A major problem with using such an array is that the captured image is severely under-sampled for at least some of the filter types. This leads to reconstructed images with strong aliasing. We make four contributions in this paper: 1) we present a comprehensive optimization method to arrive at the spatial and spectral layout of the color filter array of a GAP camera. 2) We develop a novel algorithm for reconstructing the under-sampled channels of the image while minimizing aliasing artifacts. 3) We demonstrate how the user can capture a single image and then control the tradeoff of spatial resolution to generate a variety of images, including monochrome, high dynamic range (HDR) monochrome, RGB, HDR RGB, and multispectral images. 4) Finally, the performance of our GAP camera has been verified using extensive simulations that use multispectral images of real world scenes. A large database of these multispectral images has been made available at [http://www1.cs.columbia.edu/CAVE/projects/gap\\_camera/](http://www1.cs.columbia.edu/CAVE/projects/gap_camera/) for use by the research community.

**Index Terms**—Assorted pixels, color filter array, color reproduction, demosaicing, dynamic range, multispectral imaging, signal to noise ratio, skin detection, sub-micrometer pixels.

## I. INTRODUCTION

**M**OST color image sensors use a color mosaic which is an assortment of different spectral filters. A color mosaic usually consists of three primary colors (e.g., RGB). One reason for the use of tri-chromatic filter arrays is that tri-chromatic sensing is near-sufficient in terms of colorimetric color reproducibility. It is also commonly assumed that this pixel assortment is the only practical way to sense color information with a single semiconductor image sensor.<sup>1</sup>

Recently, new image sensing technologies have emerged that use novel pixel assortments to enhance image sensing capabili-

ties. For high dynamic range (HDR) imaging, a mosaic of neutral density filters with different transmittances have been used [2], [3]. A new approach to high sensitivity imaging builds upon the standard Bayer mosaic by using panchromatic pixels [4] that collect a significantly larger proportion of incident radiation. Color filter arrays (CFAs) with more than three colors have been proposed to capture multispectral images [5], [6].

In this paper, we introduce the notion of a generalized assorted pixel (GAP) camera, which uses a mosaic with a richer assortment of filters and enables a user to produce a variety of image types from a single captured image. Each filter type in an assortment can serve to enhance a specific attribute of image quality. Examples of attributes are color reproduction, spectral resolution, dynamic range, and sensitivity. We propose a comprehensive framework for designing the spatial layout and spectral responses of the color filter array of a GAP camera. The following are the main contributions of our work: 1) We develop an optimization method to arrive at the spatial and spectral layout of the color filter array of a GAP camera. The cost function that we optimize includes terms related to colorimetric/spectral reproduction, dynamic range and signal-to-noise ratio (SNR). 2) We develop a novel algorithm for reconstructing the under-sampled channels of the image while reducing aliasing artifacts. Our approach uses a submicrometer pixel size to avoid aliasing for some of the channels. The high frequency content from these channels are then used to remove aliasing from the remaining (under-sampled) channels. 3) We have developed software that enables a user to capture a single image and then control the tradeoff of spatial resolution to generate a variety of images. The output image can be monochrome, HDR monochrome, RGB, HDR RGB, or multispectral. 4) Finally, the performance of our GAP camera has been verified using extensive simulations that use multispectral images of real world scenes. The multispectral images are used to emulate GAP camera images and results computed from the GAP images are compared with the original multispectral images. We have released a large database of high quality multispectral images (at [http://www1.cs.columbia.edu/CAVE/projects/gap\\_camera/](http://www1.cs.columbia.edu/CAVE/projects/gap_camera/)) for use by the research community.

The trend in manufacturing has been towards producing sensors with increasing numbers of smaller pixels. The time is therefore ripe for exploring more interesting pixel assortments than the ones used in the past. Furthermore, each of the previously proposed mosaics have been used to generate one specific type of output image. In contrast, our goal is to create a mosaic that lends itself to postcapture control over the output image. Since sensor fabrication is a very expensive endeavor, we have used high quality multispectral data as our ground truth

Manuscript received February 26, 2009; revised February 16, 2010. First published March 29, 2010; current version published August 18, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Oscar C. Au.

F. Yasuma, T. Mitsunaga, and D. Iso are with the Image Sensing Technology Department, Sony Corporation, 1-11-1 Osaki Shinagawa-ku, Tokyo, 141-0032, Japan (e-mail: Fumihito.Yasuma@jp.sony.com; Tomoo.Mitsunaga@jp.sony.com; Daisuke.Iso@jp.sony.com).

S. K. Nayar is with the Department of Computer Science, Columbia University, New York, New York, 10027 USA (e-mail: nayar@cs.columbia.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2010.2046811

<sup>1</sup>The Foveon X3 sensor [1] is an exception.

as well as to verify our optimized mosaic and reconstruction algorithm. Given the high quality of results we have obtained, we have begun to pursue the fabrication of a GAP sensor.

## II. BACKGROUND AND RELATED WORK

In this section, we explain the background of *optical resolution limit* using *Airy disk* and the concept of the CFA design for the smaller pixel image sensor which exceeds the *optical resolution limit*. The resolution of an optical imaging system may be limited by multiple factors, but the dominant factors are *diffraction* and *aberration*. While aberrations can be corrected for during lens design, diffraction is a fundamental limitation that cannot be avoided. Therefore, we assume an aberration-corrected optical system and focus on only diffraction. The 2-D diffraction pattern of a lens with a circular aperture is called the *Airy disk*. The width of the Airy disk determines the maximum resolution limit of the system and is given by:  $I(\theta) = I_0 \{2J_1(z)/z\}^2$ , where  $I_0$  is the intensity in the center of the diffraction pattern,  $J_1$  is the Bessel function of the first kind of order one, and  $\theta$  is the angle of observation.  $z = (\pi q/\lambda N)$ , where  $q$  is the radial distance from the optical axis in the observation plane,  $\lambda$  is the wavelength of incident light, and  $N$  is the f-number of the system. In the case of an ideal lens, this diffraction pattern is the point spread function (PSF) for an in-focus image. The Fourier transformation of the PSF is used to characterize the resolution of the optical imaging system. This quantity is referred to as the modulation transfer function (MTF). The MTF of an ideal optical system can be calculated directly from the wavelength  $\lambda$  of incident light and the f-number  $N$ , and is denoted as  $MTF_{opt}(\lambda, N) = \mathcal{F}(I(\theta))$ , where  $\mathcal{F}(\cdot)$  denotes the Fourier transformation.

Pixels typically have a rectangular shape, and their finite size contributes to the resolution characteristics of the imaging system. The MTF of an image sensor can be approximated as the Fourier transformation of a rectangular function which is denoted by  $MTF_{sensor}(p) = \mathcal{F}(s(t))$ , where

$$s(t) = \begin{cases} 1 & |t| \leq \frac{p\zeta}{2} \\ 0 & |t| > \frac{p\zeta}{2} \end{cases} \quad (1)$$

is the rectangular function,  $p$  is pixel size, and  $\zeta$  is the pixel aperture ratio, which is assumed to be 1 due to the use of on-chip microlenses. Here we define *optical resolution limit* as the maximum resolution limit of the image which is captured by the lens and sensor system. This is denoted in the frequency domain as  $MTF = MTF_{opt}(\lambda, N) \cdot MTF_{sensor}(p)$ . To compute an MTF, we use the values  $\lambda = 555$  nm (corresponding to the peak of the sensitivity of the human eye) and  $N = f/5.6$  (which is a pupil size commonly used in consumer photography). With these numbers fixed, the fundamental MTF is determined only by pixel size  $p$ . The MTFs for various pixel sizes are shown in Fig. 1. In this figure, the minimum pixel size we use is  $0.7 \mu\text{m}$ , which is the pixel size of the fabricated detector described in [7]. Note that the MTF for pixel size  $p = 0.8 \mu\text{m}$  approaches zero at about  $0.25f_s$ , where,  $f_s$  is the image sensor's sampling frequency. Thus, we can consider the optical resolution limit of an image sensor with pixel size  $p = 0.8 \mu\text{m}$  to be half of the

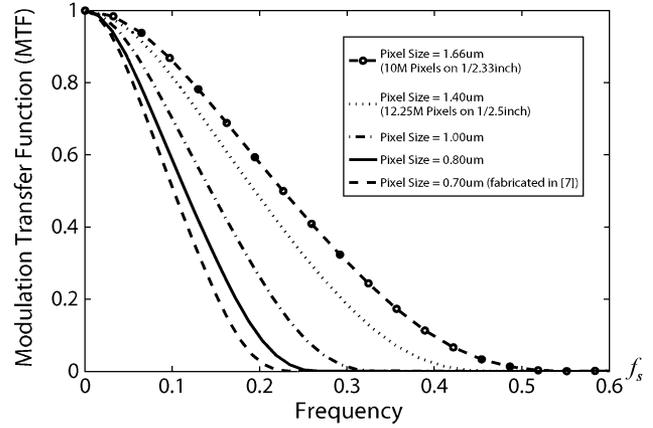


Fig. 1. Optical resolution limits (MTFs) corresponding to different pixel sizes ( $\lambda = 555$  nm and  $N = f/5.6$ ). The MTF for pixel size  $p = 0.8 \mu\text{m}$  approaches zero at about  $0.25f_s$  ( $f_s$  is the image sensor's sampling frequency). We consider the optical resolution limit of an image sensor with  $p = 0.8 \mu\text{m}$  pixel size to be half of the image sensor's Nyquist frequency. The resolution performance of a sensor with submicrometer pixels exceeds the optical resolution limit.

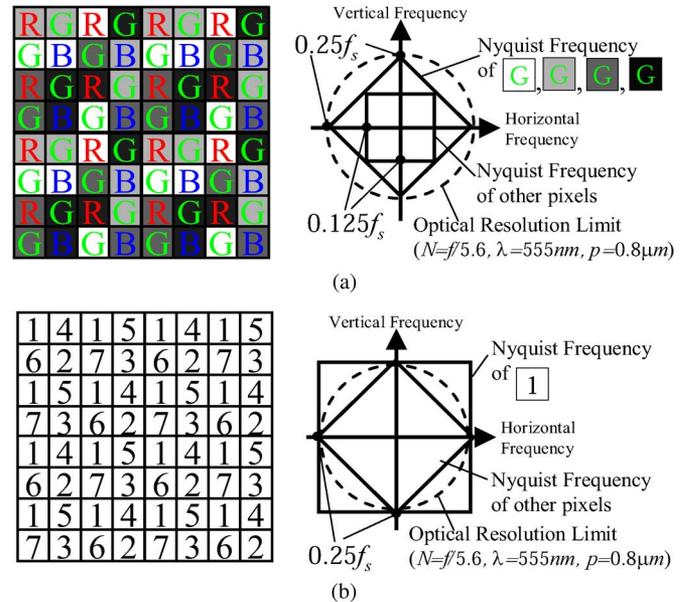


Fig. 2. Nyquist limits of previous assorted designs used with submicrometer pixel image sensors (pixel size  $p = 0.8 \mu\text{m}$ ). (a) Three colors and four exposures CFA in [3] and its Nyquist limits; (b) seven colors and one exposure CFA in [6] and its Nyquist limits.

image sensor's Nyquist frequency. From this, we can conclude that the resolution performance of a sensor with submicrometer pixels exceeds the optical resolution limit.

Fig. 2 shows the Nyquist limits when the CFA patterns of previous assorted pixels are used with the submicrometer pixel size image sensor. When the highest frequency of the input signal is lower than the Nyquist limit, aliasing does not occur, according to the sampling theorem. Therefore, aliasing is not generated at pixels marked "1" in Fig. 2(b).

Before the submicrometer image sensor is envisaged, the several advanced methods [8], [9] have been proposed for the demosaicing problem (the inverse problem of reconstructing a spatially under-sampled set whose components correspond

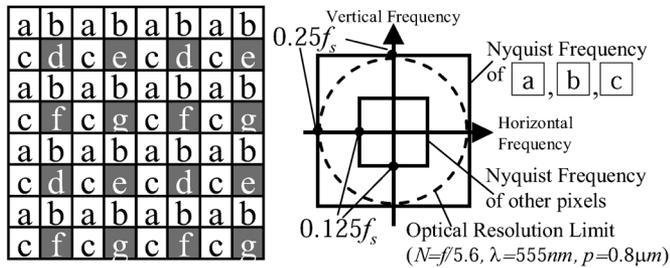


Fig. 3. Our proposed GAP mosaic (seven colors and two exposures) used with a submicrometer pixel image sensor (pixel size  $p = 0.8 \mu\text{m}$ ) and its Nyquist limits.

to particular tristimulus values). In these methods, the demosaicing problem have been solved by the frequency domain demultiplexing of a luminance component and two chrominance components. On the other hand, the CFA design method that simultaneously maximizing the spectral radii of luminance and chrominance channels is proposed in [10]. These methods are effectual to achieve the high quality RGB image from a number of under-sampled pixels. However, the CFA design for submicrometer image sensor get relief from complication to design CFA at the frequency domain, due to a lot of not under-sampled pixels. Because it became possible to arrange many aliasing-free pixels in submicrometer image sensor, we propose novel method to arrive at the spatial and spectral layout of the color filter array to provide an high quality RGB image as well as HDR and multispectral images with minimal image degradation.

### III. SPATIAL DESIGN OF GAP MOSAIC

The problem of CFA design can be simplified with the assumption that the optical resolution limit is limited to almost  $1/4$  of the sampling frequency  $f_s = 1/\Delta s$ , where  $\Delta s$  is the sampling pitch. This limit is caused by both diffraction by the lens aperture and averaging within the pixel area. In designing a CFA, our aim is to provide an aliasing-free full-color image (i.e., a full set of color triples) as well as HDR and multispectral images with minimal image degradation. To exploit this property, we propose a novel CFA to be used in conjunction with a submicrometer image sensor, which is shown in Fig. 3. This CFA consist of three primary color filters (the pixels marked “a,” “b,” and “c” in Fig. 3) and four secondary color filters (the pixels marked “d,” “e,” “f,” and “g” in Fig. 3). The number and arrangement of the primary and secondary color filters are decided as follows.

To provide an aliasing-free and full-color image, three color filters must be arranged dense enough to capture the full spatial resolution of the incident optical image. Because the incident image is band-limited to  $0.25 f_s$  (described in Section II), each of the three color filters must have a pitch of no more than 2 pixels in both the horizontal and vertical directions. This constraint is used to arrange the primary color filter (marked “a,” “b” and “c” in Fig. 3). If any of the three primary filters have a pitch of less than 2 pixels, either the remaining primary filters would be forced to have a pitch greater than 2 pixels, or there would not be sufficient space to accommodate the four secondary filters. In the former case, the CFA cannot produce

an aliasing-free and full-color image. To this end, the primary color filters are arranged as shown in Fig. 3. These aliasing-free pixels are used for reconstruction of high resolution images.

Next, we arrange the secondary filters using the remaining space on the sensor. Since the secondary filters are more sparsely arranged on the sensor, they will produce aliasing. Our approach is to remove these aliasing artifacts using the high frequencies captured (without aliasing) by the primary filters. To ensure this approach is effective, the sampling frequency of the secondary filters must be no less than  $0.125 f_s$ . To this end, each of secondary filters must have a pitch no greater than four pixels in the horizontal and vertical directions. Therefore, the number of secondary color filters can be no more than four (see Fig. 3). The accuracy of spectral reconstruction gets better as the number of basis functions increases [11]. However, previous work indicates that good reconstructions can be obtained with even seven or eight basis functions [12]. In short, the primary and secondary filters we use are reasonable to obtain better spectral reconstruction results than that supported by conventional RGB mosaic. We have validated this in our simulations and experiments. Note that we have increased the number of aliasing-free pixels from one color in conventional mosaics to three colors in our GAP design (see Fig. 2). This change results in significantly better reduction of aliasing artifacts for the secondary filters compared to previous CFAs used for multispectral imaging [6].

Due to the nature of the cost function used in our optimization procedure, the primary filters end up with spectral responses that closely resemble the red, green and blue filters commonly used in color image sensors. As a result, the primary filters can be used to compute RGB images which essentially cover the entire visible wavelength spectrum. In other words, images captured by the secondary filters, irrespective of their spectral responses, are guaranteed to be highly correlated with the images obtained using the primary filters. Consequently, the images obtained using the primary filters can be used to reduce the aliasing artifacts of the images produced by the secondary filters. Furthermore, our cost function also results in the secondary filters having lower exposures than the primary ones. Hence, by using all of the primary and secondary filters, we can obtain high dynamic range information. Finally, since the primary and secondary filters have different spectral responses, their reconstructed images can be used to obtain smooth estimates of the complete spectral reflectance distribution of each scene point, i.e., a multispectral image.

### IV. SPECTRAL RESPONSES OF GAP FILTERS

The GAP mosaic allows not only a conventional high quality RGB image but also a variety of image characteristics to be captured simultaneously. Therefore, the evaluation dimension of conventional optimization method [10] that simultaneously maximizing the spectral radii of luminance and chrominance channels is not enough for GAP concept. Monochrome and RGB images are reconstructed at high resolution from the primary filters. For HDR images, the dynamic range can be improved by using the secondary filters with a scarification of spatial resolution. Multispectral images of lower resolution can also be obtained from the secondary filters. In order to balance

these goals, we find the optimal filters for the GAP camera and design the cost function with several terms, including quality of color reproduction, reconstruction of reflectance and dynamic range.

### A. Cost Function

The value  $x_m$  measured at a pixel in the  $m$ th channel ( $m \in \{a, b, c, d, e, f, g\}$ ) is given by

$$x_m = \int_{\lambda_{\min}}^{\lambda_{\max}} i(\lambda)r(\lambda)c_m(\lambda)d\lambda \quad (2)$$

where  $i(\lambda)$  is the spectral distribution of the illumination,  $r(\lambda)$  is the spectral reflectance distribution of the scene point, and  $c_m(\lambda)$  is the spectral response of the camera's  $m$ th color channel. When the wavelength  $\lambda$  is sampled at equally-spaced  $L$  points, (2) becomes a discrete expression

$$x_m = \sum_{l=1}^L i(\lambda_l)r(\lambda_l)c_m(\lambda_l). \quad (3)$$

If we rewrite (3) in matrix form, we obtain

$$\mathbf{X} = \mathbf{C}^T \mathbf{I} \mathbf{R} \quad (4)$$

where  $\mathbf{X} = [x_a, x_b, \dots, x_g]^T$ ,  $\mathbf{C} = [c_m(\lambda_l)]$ ,  $\mathbf{I}$  is a diagonal matrix made up of the discrete illumination samples  $i(\lambda_l)$ , and  $\mathbf{R} = [r(\lambda_l)]$ . Our goal is to determine the seven spectral response functions in  $\mathbf{C}$ . The cost function includes several terms, as described in the following.

1) *Cost 1: Color Reproduction of RGB Image:* To obtain HDR RGB images, a high exposure RGB image is reconstructed from the primary filters, and a low exposure image is reconstructed from the secondary filters. The spectral responses of all the filters must ideally yield the highest color reproduction. A variety of filter rating indices have been proposed to evaluate the color reproduction characteristics of a filter [13], [14]. These indices use a cost function that minimizes the difference between the measured color of a reference material and its known color. To calculate this difference, we use the CIE 1931 XYZ color space, which is based upon direct measurements of human visual perception. The CIE  $L^*a^*b^*$  color space, which is frequently used to measure color differences, is an alternative choice. However, we chose the XYZ color space because  $L^*a^*b^*$  includes nonlinear transformations that are not ideal for evaluating HDR images. The calculation of sRGB tristimulus values (which are employed in many digital cameras and color monitors) from the CIE XYZ tristimulus values uses a linear transformation. The CIE XYZ tristimulus values are defined as  $\mathbf{Y} = \mathbf{A}^T \mathbf{I} \mathbf{R}$ , where  $\mathbf{Y}$  represents the true tristimulus values, and  $\mathbf{A}$  is a matrix of CIE XYZ color matching functions  $[\bar{x}\bar{y}\bar{z}]$ . The estimated CIE tristimulus values corresponding to the primary filters  $\hat{\mathbf{Y}}'$  can be expressed as an optimal linear transformation:  $\hat{\mathbf{Y}}' = \mathbf{T}' \mathbf{X}'$ , where  $\mathbf{X}' = [x_a, x_b, x_c]^T$ . The transformation  $\mathbf{T}'$  is determined so as to minimize the color difference:

$\min \|\mathbf{Y} - \mathbf{T}' \mathbf{X}'\|^2$ . The estimated CIE tristimulus values corresponding to the secondary filters are denoted as  $\hat{\mathbf{Y}}'' = \mathbf{T}'' \mathbf{X}''$ , where  $\mathbf{X}'' = [x_d, x_e, x_f, x_g]^T$ .

The average magnitude of the color difference between the true color  $\mathbf{Y}$  and the estimate  $\hat{\mathbf{Y}}$  over a set of  $N$  real-world objects may be used as a metric to quantify the camera's color reproduction performance. The color reproduction errors corresponding to the primary and secondary filters can therefore be written as

$$E'(\mathbf{C}) = \min_{\mathbf{T}'} \sum_{n=1}^N \|\mathbf{Y}_n - \mathbf{T}' X'_n\|^2 \quad (5)$$

$$E''(\mathbf{C}) = \min_{\mathbf{T}''} \sum_{n=1}^N \|\mathbf{Y}_n - \mathbf{T}'' X''_n\|^2. \quad (6)$$

2) *Cost 2: Reconstruction of Spectral Reflectance:* In this paper, we use the model-based spectral reconstruction method described in [15]. Fortunately, the spectral reflectance distribution of most real-world surfaces can be well-approximated using a low-parameter linear model. The linear model we use is the set of orthogonal spectral basis functions  $b_k(\lambda)$  proposed by Parkkinen *et al.* [12]

$$r(\lambda) = \sum_{k=1}^K \sigma_k b_k(\lambda) \quad (7)$$

where  $\sigma_k$  are scalar coefficients and  $K$  is the number of basis functions. By substituting (7) in (2) we get a set of equations:

$$x_m = \sum_{k=1}^K \sigma_k \int_{\lambda_{\min}}^{\lambda_{\max}} b_k(\lambda) i(\lambda) c_m(\lambda) d\lambda. \quad (8)$$

These equations can be written as  $\mathbf{X} = \mathbf{F} \cdot \boldsymbol{\sigma}$ , where  $\mathbf{F}$  is a  $M \times K$  matrix:  $\mathbf{F} = \int_{\lambda_{\min}}^{\lambda_{\max}} b_k(\lambda) i(\lambda) c_m(\lambda) d\lambda$ ,  $M$  is number of color filter channels (in our GAP mosaic,  $M = 7$ ), and  $\boldsymbol{\sigma} = [\sigma_k]$ . The spectral reflectance distribution is reconstructed by minimizing  $\|\mathbf{F} \cdot \boldsymbol{\sigma} - \mathbf{X}\|^2$ . Note that the spectral reflectance distribution of most real-world materials is known to be smooth and must be positive [15]. Thus, the reconstruction problem can be posed as a constrained minimization as follows:

$$\hat{\boldsymbol{\sigma}} = \arg \min_{\boldsymbol{\sigma}} \|\tilde{\mathbf{T}} \cdot \boldsymbol{\sigma} - \tilde{\mathbf{X}}\|^2, \quad \text{subject to } \mathbf{B} \cdot \boldsymbol{\sigma} \geq \mathbf{0} \quad (9)$$

where  $\tilde{\mathbf{F}} = [\mathbf{F}^T \alpha \mathbf{P}^T]^T$ ,  $\mathbf{P}_{lk} = \partial^2 b_k(\lambda_l) / \partial \lambda^2$  is a smoothness constraint,  $\alpha$  is a smoothness parameter,  $1 \geq l \geq L$ ,  $1 \geq k \geq K$ ,  $\tilde{\mathbf{X}} = [\mathbf{X}^T \mathbf{0}]^T$ , and  $\mathbf{B} = [b_k(\lambda_l)]$ . This regularized minimization can be solved using quadratic programming. The multispectral image's mean squared reconstruction error  $R(\mathbf{C})$  is given by

$$R(\mathbf{C}) = \sum_{n=1}^N \|\boldsymbol{\sigma}_n - \hat{\boldsymbol{\sigma}}_n\|^2 \quad (10)$$

where  $\boldsymbol{\sigma}_n$  represents the actual coefficients of the  $n$ th object and  $\hat{\boldsymbol{\sigma}}_n$  are the reconstructed coefficients. In our implementation, the

TABLE I  
OPTIMIZATION ACCURACY

|                          | Initial filters | Balanced filters<br>( $w_1 = w_2 = w_3$ ) | Emphasize RGB<br>( $w_1 > w_2 = w_3$ ) | Emphasize multispectral<br>( $w_2 > w_1 = w_3$ ) | Emphasize dynamic range<br>( $w_3 > w_1 = w_2$ ) |
|--------------------------|-----------------|---|--|--|--|
| $\Delta E'(\mathbf{C})$  | 0.0497          | 0.0429                                    | 0.0019                                 | 0.1705   | 0.0552   |
| $\Delta E''(\mathbf{C})$ | 0.0100          | 0.0055                                    | 0.0015                                 | 0.1580   | 0.0237   |
| $\Delta R(\mathbf{C})$   | 0.0624          | 0.0610                                    | 0.0623                                 | 0.0478   | 0.0641   |
| $DR_{GAP}$               | 58.2970         | 62.9213                                   | 57.3209                                | 58.8955  | 78.7337  |
| $SNR_{GAP}$              | 34.7069         | 32.3694                                   | 35.1988                                | 34.4051  | 24.1577  |

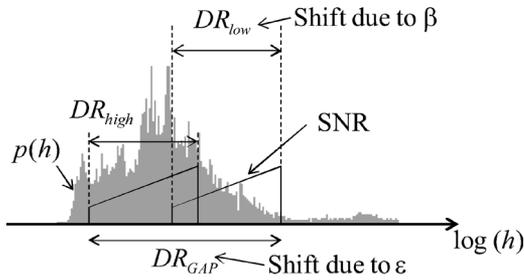


Fig. 4. Histogram  $p(h)$  of image irradiance and the dynamic range of the GAP camera. The extended dynamic range  $DR_{low}$  can be shifted using  $\beta$  and the total dynamic range  $DR_{GAP}$  can be shifted using  $\epsilon$ .

number of basis functions  $K$  is 8 and the smoothness parameter  $\alpha$  is set to 64.0 [15].

3) *Cost 3: Dynamic Range and SNR*: The third criterion of GAP filters is that they should maximize the dynamic range while keeping SNR as large as possible. To achieve HDR imaging, our secondary filters have lower transmittances than the primary filters, as mentioned earlier. This may cause deterioration of the signal-to-noise ratio (SNR) for the secondary filters. This tradeoff can be controlled based upon the ratio of the exposures of the primary and secondary filters:  $\beta = (e_{max}/e_{min})$ , where  $e_{max}$  is the average exposure of the primary filters and  $e_{min}$  is the average exposure of the secondary filters. Therefore,  $\beta$  is determined by  $\mathbf{C}$  in (4). Our goal here is to determine the value of  $\beta$  that best balances extension of dynamic range versus reduction of SNR. We wish to choose a value for  $\beta$  which minimizes the number of the over-exposed and under-exposed pixels and maximizes the SNR of well-exposed pixels. This is identical to maximizing the integral of the SNR of the well-exposed pixels over the histogram  $p(h)$  of image irradiance (see Fig. 4). The amount of light incident on the detector is controlled by camera exposure  $\epsilon$ , which is determined by the shutter speed and pupil size. When the detector is linear, the signal is proportional to the incident light:  $V = \epsilon h$ . The signal of secondary filter can be denoted as  $V_{low} = \epsilon h/\beta$ . We assume the noise arisen from the light incident is shot noise, therefore, the SNR is  $V/\sqrt{V}$  [16]. Thus, the problem can be posed as the following maximization:

$$(\hat{\beta}, \hat{\epsilon}) = \arg \max_{(\beta, \epsilon)} \left( \int \frac{V}{\sqrt{V}} p(h) dh + \int \frac{V_{low}}{\sqrt{V_{low}}} p(h) dh \right) \quad (11)$$

where,  $\hat{\beta}$  and  $\hat{\epsilon}$  are the optimal values of  $\beta$  and  $\epsilon$ .

To solve the previous optimization, we approximate the SNR of under-exposed and over-exposed pixels with zero

$$(\hat{\beta}, \hat{\epsilon}) = \arg \max_{(\beta, \epsilon)} \sum_{o=1}^O \left( \sum_{h=V_{min}/\epsilon}^{V_{full}/\epsilon} \frac{V}{\sqrt{V}} p_o(h) + \sum_{h=\beta V_{min}/\epsilon}^{\beta V_{full}/\epsilon} \frac{V_{low}}{\sqrt{V_{low}}} p_o(h) \right) \quad (12)$$

where  $p_o(h)$  is  $o$ th image's output histogram,  $V_{full}$  represents the full-well capacity of the detector,  $V_{min}$  is the minimum output of the detector. The error in the computed high dynamic range image is defined as

$$D(\mathbf{C}) = (\beta - \hat{\beta})^2. \quad (13)$$

4) *Total Cost Function*: We confirmed that the each of the previously mentioned cost functions is convergent and maximize an each image quality (see Table I). Thus, to achieve the balanced quality image our final cost function becomes a weighted sum of the individual costs

$$G = w_1 \{E' + E''\} + w_2 R + w_3 D. \quad (14)$$

The weights  $(w_1, w_2, w_3)$  are determined according to the image quality requirements of the application for which the GAP camera is manufactured. Since all camera filters must have positive spectral responses ( $\mathbf{C}$  must be positive), the optimization of  $\mathbf{C}$  can be written as

$$\mathbf{C} = \arg \min_{\mathbf{C}} G, \quad \text{subject to } \mathbf{C} \geq \mathbf{0}. \quad (15)$$

5) *Initial Guesses for Filter Spectral Responses*: Note that in the absence of additional constraints, our goal of finding the seven spectral response functions in  $\mathbf{C}$  is an intractable optimization problem. Therefore, we assign initial guesses to the filter responses. These filter guesses are driven by two factors: 1) they are selected from a set of 177 commercially available optical band pass filters [17] and on-chip filters [1]; and 2) the commercial filters are assigned to the seven channels based upon only one of our cost functions, namely, color reproduction. That is, we find the primary filters  $\mathbf{C}'_0$  and secondary filters  $\mathbf{C}''_0$  such that

$$\min_{\mathbf{C}'_0} E(\mathbf{C}'_0) \quad (\mathbf{C}'_0 \in \mathbf{C}_0) \quad (16)$$

$$\min_{\mathbf{C}''_0} E(\mathbf{C}''_0) \quad (\mathbf{C}''_0 \in \mathbf{C}_0) \quad (17)$$

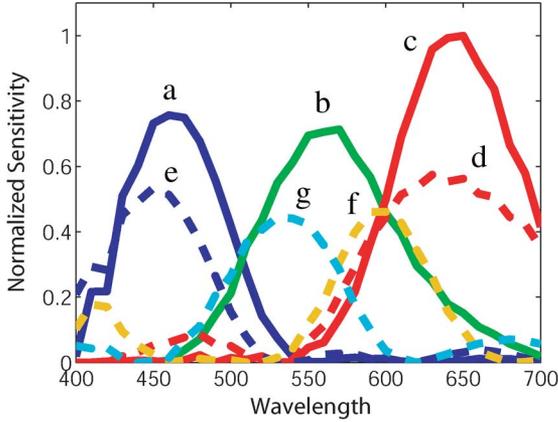


Fig. 5. Spectral responses of the seven optimized filters. The secondary filters (“d,” “e,” “f,” “g”) have lower exposures than the primary ones (“a,” “b,” “c”). Hence, using the primary and secondary filters, we can obtain high dynamic range information. Since each filter has different spectral response, their reconstructed images can also be used to obtain smooth estimates of the complete spectral reflectance distribution.

where  $C_0$  is the set of 177 commercial filters. Once the seven assignments are made in this way, they are used as initial guesses in the final stage of the optimization. This final stage is a constrained nonlinear minimization of (15) which requires the use of an iterative algorithm. In our implementation, we used the `fmincon` routine of Matlab. For the weights, we have used  $w_1 = w_2 = w_3 = 1.0$ . As mentioned earlier, these weights can be chosen differently to meet the needs of the application.

### B. Results of GAP Filter Optimization

Using the previously shown optimization, we obtain the optimal filter spectra shown in Fig. 5. We use the spectral reflectance distribution of the color patches in the Macbeth color chart and the Munsell color book as the known references  $r(\lambda)$ , the illuminance spectrum of D65 for  $i(\lambda)$ , and a high dynamic range image database [18], [19] for computing the histograms  $p(h)$ . Three observations are worth making. First, as a result of the color reproduction term in the cost function, the primary filters are close in their responses to red, green and blue filters. Second, due to the spectral reconstruction term, the computed filters nicely sample the visible spectrum, which enables the GAP camera to produce reliable multispectral images. Third, due to the HDR and SNR term, the primary filters have higher transmittances than the secondary filters. Dynamic range is often defined as  $DR = 20 \log_{10}(V_{\text{full}}/V_{\text{min}})$ , where  $V_{\text{full}}$  represents the full-well capacity of the detector, and  $V_{\text{min}}$  is the minimum output of the detector. In the case of a GAP camera,  $V_{\text{min}}$  is fixed, but the maximum detectable level becomes  $\beta V_{\text{full}}$  [2]. Hence, the dynamic range of a GAP camera is

$$DR_{GAP} = 20 \log_{10} \frac{\beta V_{\text{full}}}{V_{\text{min}}}. \quad (18)$$

The SNR can be written as:  $SNR = 20 \log_{10}(V/N)$ , where  $V$  is the signal and  $N$  is the noise. In this paper, the noise of the detector is defined as  $N = \sqrt{N_{\text{shot}}^2 + N_{\text{dark}}^2}$ , where  $N_{\text{shot}} = \sqrt{V}$  is the shot noise,  $V$  is the signal, and  $N_{\text{dark}}$  is the dark noise [16]. The signal corresponding to a secondary filter can be expressed using the exposure ratio  $\beta$  as  $V'' = V'/\beta$ , where

$V'$  is a signal due to a primary filter. When the signal due to the primary filter is not saturated, the signal due to the secondary filter can be determined from the primary signal. The SNR for a secondary filter when the primary signal is saturated is the worst-case SNR of the GAP mosaic

$$SNR_{GAP} = 20 \log_{10} \frac{V_{\text{full}}/\beta}{N_{\text{max}}} \quad (19)$$

where  $N_{\text{max}} = \sqrt{N_{\text{shot}}'^2 + N_{\text{dark}}^2}$ , and  $N_{\text{shot}}' = \sqrt{V_{\text{full}}/\beta}$ . In our calculation, we have used  $V_{\text{full}} = 3500e^-$  and  $N_{\text{dark}} = 33e^-$  (see [7]).

Table I shows the errors in the color reproduction and spectral reconstruction components of our cost function  $G$ , the estimated dynamic range, and the SNR of the initial and final (optimized) set of seven filters. Note that all the evaluated value except SNR are reduced as a result of the optimization. The deterioration of SNR is kept low at around 2.3 dB while the dynamic range is improved by about 4.6 dB. If we use a large value for the weight corresponding to one of the image types (RGB, multispectral or HDR), the reconstruction quality for that image type improves while the quality of the other image types deteriorates. This can be seen from the results in Table I. When all the weights are chosen to be equal, we achieve a more balanced performance—all the image types are reconstructed with reasonable quality.

### V. POST-CAPTURE CONTROL OF IMAGE TYPES

At each pixel of the GAP mosaic in Fig. 3, there is only one color measurement, which means that the other colors must be estimated from neighboring pixels in order to produce interpolated output images (irrespective of their type). This process is commonly referred to as “demaicing.”

Denoting  $\Lambda_m$  as the set of pixel locations  $(i, j)$  for filter  $m \in \{a, b, c, d, e, f, g\}$ , a mask function for each filter can be defined as

$$W_m(i, j) = \begin{cases} 1, & (i, j) \in \Lambda_m \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

In the GAP mosaic, there are seven types of color channels:  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $e$ ,  $f$ , and  $g$ . Therefore, the observed data  $y(i, j)$  is

$$y(i, j) = \sum_{m \in \{a, b, c, d, e, f, g\}} W_m(i, j) x_m(i, j) \quad (21)$$

where  $x_m$  is  $m$ th channel’s full resolution image, given by (2). Fig. 6 shows the complete framework of our proposed multi-modal image reconstruction. The interpolated image after demosaicing is denoted as  $\hat{x}$ . Different types of images are reconstructed from all interpolated images by simply changing the image reconstruction matrix

$$I_p = Z_p \mathbf{A} \quad (22)$$

where  $I_p$  is the reconstructed image (which can be monochrome, HDR monochrome, RGB, HDR RGB, or multispectral),  $Z_p$  is an image reconstruction matrix, and  $\mathbf{A}$  is an interpolated image set denoted as a vector:  $\mathbf{A} = [\hat{x}_a \hat{x}_b \hat{x}_c \hat{x}_d \hat{x}_e \hat{x}_f \hat{x}_g \hat{x}_{LEM}]^T$ , where  $\hat{x}_{LEM}$  is the interpolated low-exposure monochrome image. The user

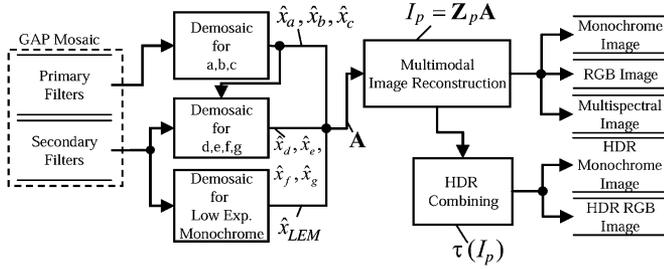


Fig. 6. Overview of the proposed multimodal image reconstruction.

can control the tradeoff of spatial resolution to generate a variety of images by changing the image reconstruction matrix  $Z_p$  from a single captured image  $y(i, j)$ . We now describe the different processing operations of Fig. 6.

#### A. Demosaicing for “a,” “b,” and “c” Images

As described in Section III, images captured by the primary filters do not suffer from aliasing. Therefore, we can estimate the missing data using a simple interpolation. The “a,” “b,” and “c” channels’ images  $\hat{x}_a$ ,  $\hat{x}_b$ , and  $\hat{x}_c$  are reconstructed using just the data measured by the primary filters, to maintain high resolution. For interpolation we use a Finite Impulse Response (FIR) filter  $F(i, j)$

$$\hat{x}_\nu(i, j) = W_\nu(i, j)y(i, j) + \bar{W}_\nu(i, j)[F(i, j) * y(i, j)] \quad (23)$$

where  $\nu = a, b$ , or  $c$ ,  $*$  denotes convolution, and  $\bar{W}(i, j) = 1 - W(i, j)$ . To minimize the loss of high frequencies due to interpolation, we used Matlab’s `fir2` function to find a FIR filter of size  $30 \times 30$  that passes all frequencies. This FIR filter is a product of two orthogonal 1-D sinc functions with a cutoff at  $0.25 f_s$ .

#### B. Demosaicing for “d,” “e,” “f,” and “g” Images

Interpolated secondary filter images— $\hat{x}_d$ ,  $\hat{x}_e$ ,  $\hat{x}_f$ , and  $\hat{x}_g$ —can be computed using only the “d,” “e,” “f,” or “g” pixels. However, this results in severe aliasing [see Fig. 7(a)]. In conventional demosaicing methods for RGB mosaics [20], an assumption of strong positive interchannel correlation (the color ratios within an image segment are assumed to be constant) is commonly used so as to suppress aliasing of the sparsely sampled channels (R, B) by estimating the amount of aliasing from the high frequency information of a densely sampled channel (G). (This method can be also discussed with the luminance component and the chrominance component at the frequency domain [8], [9]). However, this assumption often results in artifacts because the differences in the spectral responses of RGB filters cause the interchannel correlation of RGB to be not always strongly positive. On the other hand, our aliasing reduction method can exploit the inherent interchannel correlations within GAP mosaic. As shown in Fig. 5, one primary filter color can be chosen for each secondary filter color in terms of similarity of the spectral response, with high expectation of strong positive interchannel correlation due to strong overlap between the spectral responses of the chosen primary and secondary channels. For example, channel “a” is chosen as a strongly correlated channel with channel “e.” So we

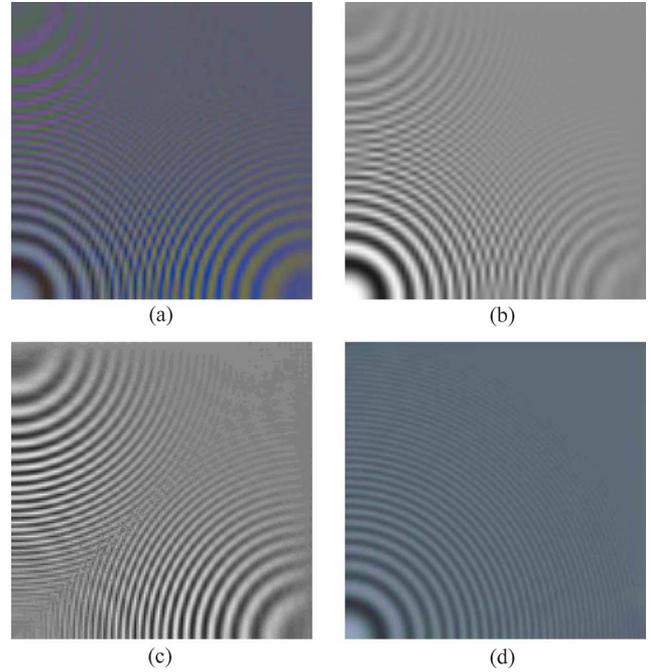


Fig. 7. Aliasing reduction algorithm (simulated with  $N = f/5.6$ ,  $p = 0.8 \mu\text{m}$ ). (a) Low exposure RGB image computed from the secondary filters without aliasing reduction (false color artifacts caused by aliasing are observed). (b) Downsampled image  $\Omega * \{W_e(i, j)\hat{x}_a(i, j)\}$  computed using the pixels with primary filter “a.” (c) Aliasing  $\Upsilon_e(i, j)$  estimated using (b) and the full resolution image for channel “a” (brightness enhanced for visualization). (d) Low exposure RGB image obtained after aliasing reduction using the estimated aliasing in (c).

first sample the interpolated full resolution “a” filter image  $\hat{x}_a$  at all “e” locations to estimate the aliasing of “e” filter image. These samples are then used to compute a full resolution image for the “e” filter:  $\Omega * \{W_e(i, j)\hat{x}_a(i, j)\}$ , where  $\Omega$  represents an low-pass filter. We used a bilinear interpolation filter for  $\Omega$ . Aliasing can be inferred by subtracting the original  $\hat{x}_a$  image from this interpolated one. To get the final estimate of aliasing in the “e” channel,  $\Upsilon_e$ , we used assumption that the color ratios within an object in an image is constant. The color ratio  $\Psi_{ae}$  within an object in “a” and “e” channel image is used for the interpolation of the “e” channel. The estimated aliasing  $\Upsilon_e$  in the “e” channel is given by

$$\Upsilon_e(i, j) = \Psi_{ae}(i, j) \cdot [\Omega * \{W_e(i, j)\hat{x}_a(i, j)\} - \hat{x}_a(i, j)] \quad (24)$$

where  $\Psi_{ae}(i, j)$  is

$$\Psi_{ae}(i, j) = \frac{\Omega * \{W_e(i, j)y(i, j)\}}{\Omega * \{W_e(i, j)\hat{x}_a(i, j)\}} \quad (25)$$

The image  $\hat{x}_e$  with reduced aliasing is obtained as

$$\hat{x}_e(i, j) = \Omega * \{W_e(i, j)y(i, j)\} - \Upsilon_e(i, j). \quad (26)$$

Since the same sampling  $W_e$  and low-pass filter  $\Omega$  are used, the aliasing component in  $\Omega * \{W_e(i, j)y(i, j)\}$  and estimated aliasing  $\Upsilon_e(i, j)$  are identical. Our aliasing estimation technique also assumes positive interchannel correlation. It is not effective in the case of negative interchannel correlation, as is the case with previous techniques. Therefore, we select our channel pairs

for aliasing reduction such that they have the highest positive correlation between them. Since the output signal of the sensor is a product of the incident spectral distribution and the sensor's spectral response, the channels with larger overlap in spectral response tend to have higher positive correlation. The other secondary filter images— $\hat{x}_d, \hat{x}_f, \hat{x}_g$ —can be similarly computed. We select the color filter pairs of  $\hat{x}_b$  with  $\hat{x}_f$ ,  $\hat{x}_b$  with  $\hat{x}_g$  and  $\hat{x}_c$  with  $\hat{x}_d$ , because that interchannel correlation becomes stronger positive due to maximizing the spectral overlap. Fig. 7 shows an example that illustrates the efficacy of this aliasing reduction technique.

### C. Demosaicing for Low Exposure Monochrome Images

In order to compute an HDR monochrome image, we need to first compute a low exposure monochrome image. We can construct this low exposure image using only the four secondary filters which have lower exposure and also collectively cover the entire visible spectrum (Fig. 5). In Fig. 3, we see that four different secondary pixels are arranged diagonally about each “a” pixel. Therefore, the monochrome value at each “a” pixel can be computed as the average of the measurements at the four neighboring secondary pixels:  $L(i, j) = W_a(i, j)\{\mathbf{Q}_D * y(i, j)\}$ , where

$$\mathbf{Q}_D = \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} \end{pmatrix}. \quad (27)$$

Note that by adding four pixels in a diagonal neighborhood, aliasing caused by half-pixel phase shifts gets canceled out.<sup>2</sup> The values at the “a” pixels are then interpolated for all other pixels to yield the low exposure monochrome image  $\hat{x}_{LEM}$

$$\hat{x}_{LEM}(i, j) = L(i, j) + W_s \{\mathbf{Q}_D * L(i, j)\} + W_b \{\mathbf{Q}_H * L(i, j)\} + W_c \{\mathbf{Q}_V * L(i, j)\} \quad (28)$$

where  $W_s(i, j) = \begin{cases} 1, & (i, j) \in \{d, e, f, g\} \\ 0, & \text{otherwise} \end{cases}$ , and  $\mathbf{Q}_H = \mathbf{Q}_V^T = [0.5 \ 0 \ 0.5]$ .

### D. Multimodal Image Reconstruction

As shown in Fig. 3, the primary filters capture images at a higher sampling frequency than the secondary filters. Thus, the spatial resolution of  $\hat{x}_a, \hat{x}_b$ , and  $\hat{x}_c$  is higher than that of  $\hat{x}_d, \hat{x}_e, \hat{x}_f$ , and  $\hat{x}_g$ . Although the aliasing of images reconstructed from secondary filters is reduced due to our aliasing reduction process, the usage of  $\hat{x}_{LEM}$  and  $\hat{x}_d, \hat{x}_e, \hat{x}_f, \hat{x}_g$  slightly degrades the spatial resolution of the reconstructed image. We now describe how each of the different output images can be reconstructed with the least loss in spatial resolution.

1) *Reconstruction of Monochrome Image:* Monochrome image  $I_M$  is reconstructed using just the data measured by the primary filters to maintain high resolution

$$I_M = \mathbf{Z}_M \mathbf{A} \quad (29)$$

<sup>2</sup>Note that this aliasing reduction method can only be used for a monochrome image computed from the low exposure secondary filters and not for a color image computed from the same.

where  $\mathbf{Z}_M = [\kappa_a \ \kappa_b \ \kappa_c \ 0 \ 0 \ 0 \ 0 \ 0]$  where  $\kappa_a, \kappa_b$  and  $\kappa_c$  are the coefficients of the transformation from the outputs of the primary filters to monochrome image. We used  $\kappa_a = \kappa_b = \kappa_c = 1$ , because the weights  $\kappa_a, \kappa_b$  and  $\kappa_c$  that define monochrome should be selected to optimize spatial acuity and not to match as closely as possible the human luminosity function.

2) *Reconstruction of RGB Image:* To construct the RGB image  $I_{RGB}$ , we use the color reproduction matrix  $\mathbf{T}'$  (Section IV-A-1) and  $\mathbf{H}$  (linear transformation from CIE XYZ to sRGB) to combine the information in the  $\hat{x}_a, \hat{x}_b$ , and  $\hat{x}_c$  images computed using only the primary pixels [(23)]

$$I_{RGB} = \mathbf{Z}_{RGB} \mathbf{A} \quad (30)$$

$$\text{where } \mathbf{Z}_{RGB} = \begin{bmatrix} \mathbf{HT}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

3) *Reconstruction of HDR Monochrome and HDR RGB Image:* We combine the monochrome image  $I_M$  and the low exposure monochrome image  $\hat{x}_{LEM}$  (to which aliasing reduction has been applied) to produce the HDR monochrome image  $I_{HDRM}$

$$I_{HDRM} = \tau(\mathbf{Z}_{HDRM} \mathbf{A}) \quad (31)$$

where  $\mathbf{Z}_{HDRM} = \begin{bmatrix} \kappa_a & \kappa_b & \kappa_c & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ , and  $\tau(\cdot)$  is the processing that combining the HDR image from different exposure images, and is based upon the method described in [2].

Similarly, we obtain the HDR RGB image  $I_{HDRRGB}$  from the RGB image  $I_{RGB}$  and the low exposure RGB image that is obtained by multiplying the secondary filter images by a color reproduction matrix  $\mathbf{T}''$  and color space conversion  $\mathbf{H}$

$$I_{HDRRGB} = \tau(\mathbf{Z}_{HDRRGB} \mathbf{A}) \quad (32)$$

$$\text{where } \mathbf{Z}_{HDRRGB} = \begin{bmatrix} \mathbf{HT}' & \mathbf{0} \\ \mathbf{0} & \mathbf{HT}'' \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

4) *Reconstruction of Multispectral Image:* For multispectral imaging,  $\hat{x}_a, \hat{x}_b, \hat{x}_c$  and the  $\hat{x}_d, \hat{x}_e, \hat{x}_f, \hat{x}_g$  images (to which aliasing reduction has been applied) are used to reconstruct the spectral reflectance distribution of an object using the method given by (9)

$$\mathbf{X} = \mathbf{Z}_{MS} \mathbf{A} \quad (33)$$

where  $\mathbf{Z}_{MS} = \begin{bmatrix} \mathbf{I}_{7 \times 7} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ , and  $\mathbf{I}_{7 \times 7}$  is 7 by 7 identity matrix.

## VI. COMPARISON WITH OTHER MOSAICS

The performance of demosaicing methods depends upon the spatial layout and spectral responses of the color filters used, both of which vary from detector to detector. Moreover, previous CFA mosaics [3], [6] were not designed for controlling the tradeoff of spatial resolution to generate a variety of images. Therefore, a direct comparison of image qualities is difficult to perform. Instead, Table II shows a qualitative comparison between the performances of our GAP mosaic, a previously proposed assorted pixel CFA [3] and a previously proposed mul-

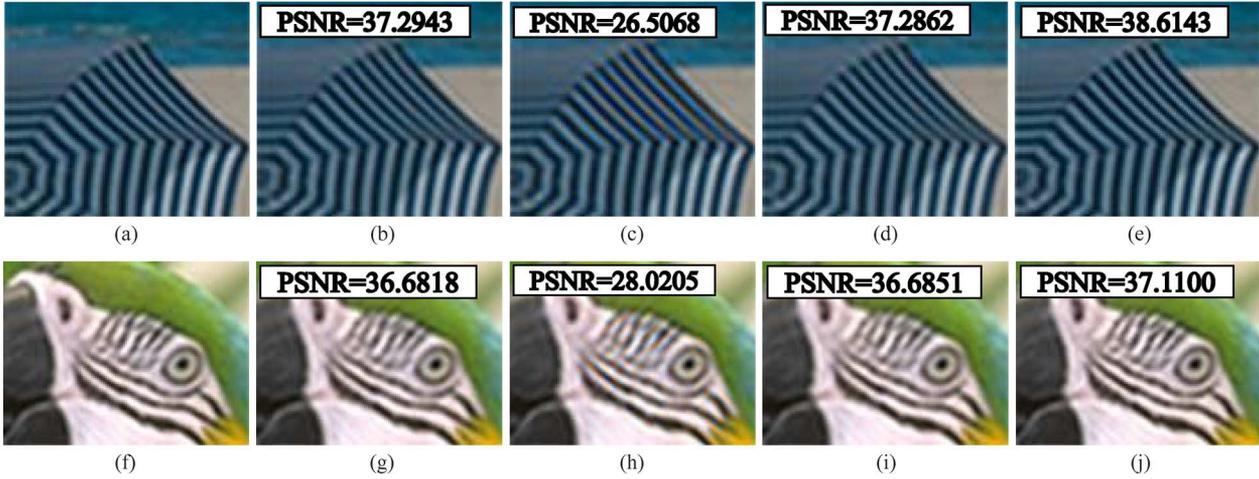


Fig. 8. Comparison of RGB images. (a), (f) Ground truth (simulated with  $N = f/5.6, p = 0.8 \mu\text{m}$ ). (b), (g) Assorted Pixels [3]. (c), (h) Multispectral CFA [6]. Aliasing artifacts are observed over the stripe pattern in (c) and (h). (d), (i) GAP Camera. (e), (j) BAYER CFA demosaiced with AHD [21].

TABLE II  
COMPARISON OF THE GAP CAMERA WITH PREVIOUS ASSORTED PIXELS [3] AND MULTISPECTRAL CFA [6]. THE SHADED CFA OFFERS THE BEST QUALITY IMAGE FOR EACH IMAGE TYPE

| Mosaic Pattern           | Monochrome                      | RGB | HDR Monochrome | HDR RGB | Multispectral |
|--------------------------|---------------------------------|-----|----------------|---------|---------------|
| <p>Assorted Pixel</p>    | <p>Optical Resolution Limit</p> |     |                |         | N/A           |
| <p>Multispectral CFA</p> |                                 |     | N/A            | N/A     |               |
| <p>GAP Camera</p>        |                                 |     |                |         |               |

tispectral CFA [6]. Note that when the difference of exposures for HDR imaging is disregarded in the case of assorted pixels, it is identical to the Bayer mosaic [20]. For monochrome images, although there is fundamentally no spatial resolution difference between the three mosaics, the monochrome image of the GAP mosaic is reconstructed from the “a,” “b,” and “c” channels, which together cover all visible wavelengths (see Fig. 5). Therefore, the GAP mosaic can reproduce monochrome images more accurately than other CFAs. In Fig. 8, we compare the RGB images which are computed using three CFAs (assorted pixels, multispectral CFA and GAP). We use natural images captured using Kodak PhotoCD (used in [22]). The same demosaicing al-

gorithm (described in Section V-A) was used for all three CFAs. We also compare these RGB images with a Bayer mosaic with the up-to-dated demosaicing algorithm [21] in Fig. 8. We can see aliasing artifacts in the images produced by the multispectral CFA. This is because that the R and B filters of that CFA are not dense enough (see Table II). We evaluate the degradation by aliasing artifacts using peak signal-to-noise ratio (PSNR) in Fig. 8. The PSNR of multispectral CFA is deteriorated about 10 dB due to aliasing artifacts. The difference of PSNR between images provided from GAP and Bayer mosaics is about 1 dB, so that the GAP mosaic can reproduce RGB image nearly as same quality as Bayer mosaic. Note that the GAP mosaic can

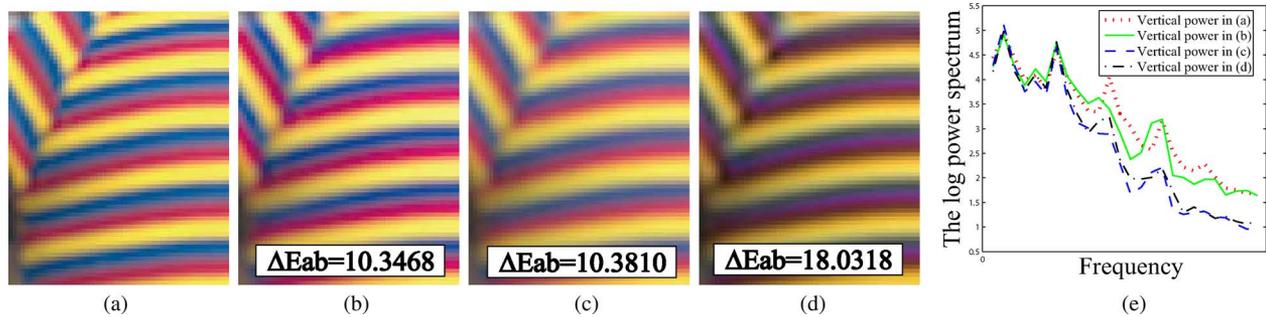


Fig. 9. Comparison of aliasing reductions. These RGB images were simulated with seven spectral filters (Fig. 5) and multispectral images (Section VII-B). (a) Ground truth (simulated with  $N = f/5.6$ ,  $p = 0.8 \mu\text{m}$  and the CIE 1931 RGB Color matching functions). (b) Reconstructed from the primary filters (“a,” “b,” and “c”) of GAP Camera. (c) Reconstructed from the secondary filters (“d,” “e,” “f” and “g”) of GAP Camera (brightness enhanced for visualization). The aliasing artifacts are reduced by using (b) that aliasing-free image. (d) Reconstructed from the multispectral CFA. The green spectrum (“b” in Fig. 5) is used as the most dense filter arrangement (“1” in Fig. 2). False color artifacts are observed in (d). (e) The vertical components of the logarithmic Fourier power spectrum of RGB images. The high frequency component of image (b) close to ground truth (a), although degradation of resolution can be observed in (c) and (d).

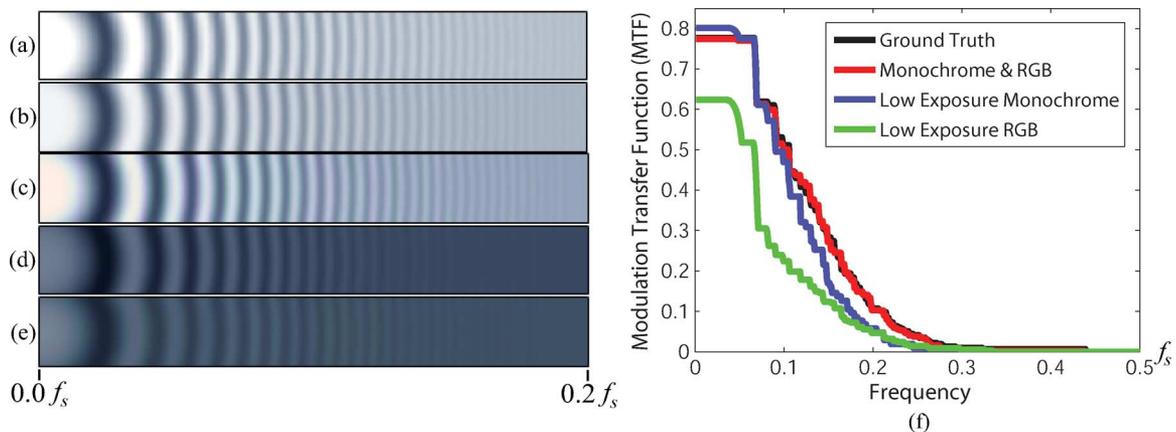


Fig. 10. Result of multimodal demosaicing for a CZP chart. (a) Ground truth (simulated with  $N = f/5.6$ ,  $p = 0.8 \mu\text{m}$ ). (b) Demosaiced monochrome image. (c) Demosaiced RGB. (d) Demosaiced (with aliasing reduction) low-exposure monochrome image. (e) Demosaiced (with aliasing reduction) low exposure RGB image. (f) MTFs of ground truth and demosaiced CZP images.

provide not only the high quality RGB images (on par with the Bayer mosaic and assorted pixels) but also HDR RGB and multispectral images. The GAP mosaic can also create HDR RGB images of the same resolution as the assorted pixel array. However, because the assorted pixel array has four different exposures, it is more effective at extending dynamic range than the GAP camera. The aliasing artifacts are caused by under-sampling, and are reduced by using aliasing-free (densely-sampled) channels of the CFA. Since our aliasing estimation technique assumes positive interchannel correlation (as in previous techniques [20]), negative interchannel correlation gives rise to false color artifacts [see Fig. 9(d)]. Our GAP mosaic is less likely to produce such artifacts compared to previous CFAs (the reasons are given in Section V-B). Fig. 9 compares aliasing reduction results of GAP with the multispectral CFA. In the case of the multispectral CFA, the aliasing artifacts of under-sampled pixels can be reduced using only one aliasing-free channel [“1” in Fig. 2(b)]. Thus, the color difference  $\Delta E_{ab}$  of multispectral CFA is increased. The image reconstructed using GAP does not include false color artifacts, although color reproduction performance is not perfect. In Fig. 9(e), the high frequency component of GAP image reconstructed from primary filters close to ground truth. This is one of GAP’s advantages—since all three primary channels have sufficient density, the computed RGB

images are artifact free. In summary, when high spatial resolution is necessary, the GAP camera offers images with quality that is similar to, or better than, other CFA mosaics.

## VII. EXPERIMENTAL RESULTS

### A. Results for the CZP Chart

Fig. 10(a) shows a synthesized circular zone plate (CZP) image computed using a diffraction-limited model of a lens with an f-number of 5.6 and  $0.8 \mu\text{m}$  pixel size (without considering noise). This serves as the ground truth. Fig. 10(b)–(e) show demosaiced images computed from a GAP mosaic image—(b) monochrome, (c) RGB, (d) low exposure monochrome, and (e) low exposure RGB. Fig. 10(f) shows MTFs of these demosaiced images. The monochrome and RGB images computed using the primary filters are very close to the ground truth. The low exposure monochrome image has an MTF of 0.1 at  $0.1754 f_s$ , while the low exposure RGB image’s MTF is 0.1 at  $0.1647 f_s$ . For standard monochrome and RGB images this occurs at  $0.2125 f_s$ . This demonstrates that our GAP mosaic with multimodal demosaicing allows a user to control the tradeoff between spatial resolution and radiometric details of the output image.

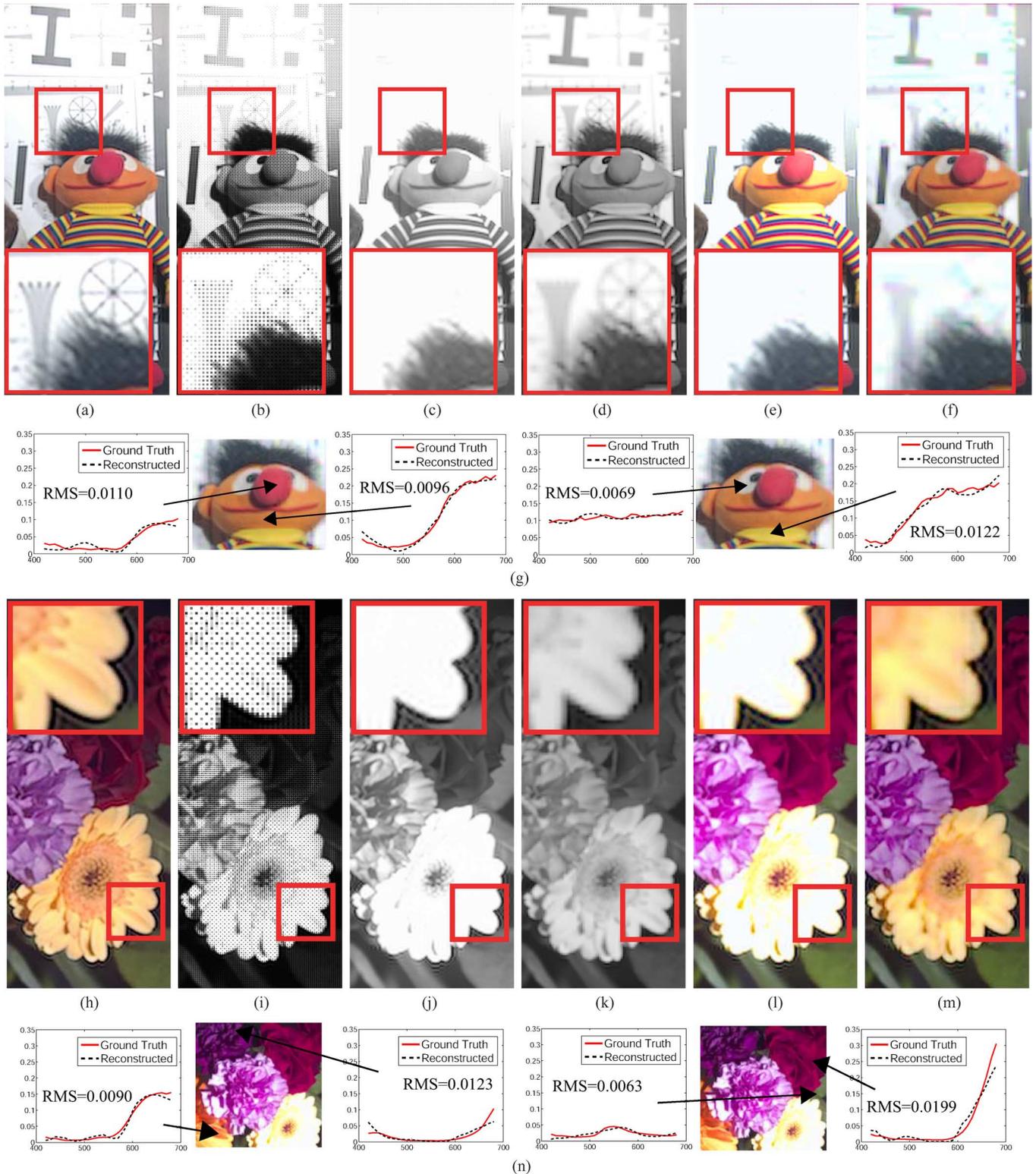


Fig. 11. Results for real scenes. (a), (h) Ground truths (simulated with  $N = f/5.6$ ,  $p = 0.8 \mu\text{m}$ ). (b), (i) GAP mosaic (raw) images. (c), (j) demosaiced mono-chrome images. (d), (k) HDR mono-chrome images. (e), (l) RGB images. (f), (m) HDR RGB images. (g), (n) multispectral images and examples of reconstructed spectral reflectance distributions.

*B. Experiments With Multispectral Images*

We also captured 31-band multispectral images (400–700 nm, at 10 nm intervals) of several static scenes using a tunable filter

(VariSpec Liquid Crystal Tunable Filter) and a cooled CCD camera (Apogee Alta U260,  $512 \times 512$  pixels). We have captured multispectral images for a wide variety of objects and materials, including, textiles, skin, hair, real and fake fruits and vegetables,



Fig. 12. Results for real scenes. (a) Ground truth (simulated with  $N = f/5.6$ ,  $p = 0.8 \mu\text{m}$ ). (b) GAP mosaic (raw) image. (c) demosaiced monochrome image. (d) HDR monochrome image. (e) RGB image. (f) HDR RGB image. (g) Multispectral image and reconstructed spectral reflectance distributions. (h) Result of skin detection for RGB Image by using simple correlation-based method. (i) Result of skin detection applied to a multispectral image computed from the GAP image.

candy, drinks, paints, etc. We believe this database could be valuable to researchers working in areas related to multispectral imaging. The database has been made publicly available at: [http://www1.cs.columbia.edu/CAVE/projects/gap\\_camera/](http://www1.cs.columbia.edu/CAVE/projects/gap_camera/).

The multispectral images were used to simulate images captured with a GAP mosaic. Fig. 11 and 12 shows these as well as our multimodal demosaicing results for two different scenes. For both scenes, the textures and colors of saturated regions in the monochrome and RGB images become visible in the corresponding HDR images. As expected, one can see more details in the HDR monochrome images than in the HDR RGB images. We also experimented within skin detection using RGB image and multispectral data. Fig. 12(h) shows the result of skin detection using an RGB image by using a simple correlation-based method. Fig. 12(i) shows the result of skin detection applied to multispectral images computed from GAP images [see Fig. 12(g)]. Note that the scene shown in Fig. 12 includes a real face (skin) on the right and a photo of the same face (printed paper) on the left. As seen in Fig. 12(h), these two faces (real and fake) are difficult to distinguish using the RGB image—skin detection based upon color analysis finds both the faces although only one of them is real. In contrast, skin detection applied to the multispectral image computed from the GAP image results in the desired result—only the real face is found as pixels within it have the spectrum of real skin [see Fig. 12(i)].

## VIII. CONCLUSION

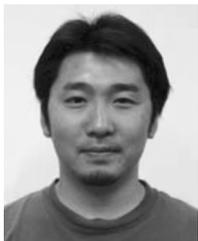
In this paper, the concept of a generalized assorted pixel camera has been presented. We have developed a general framework for designing GAP cameras that can simultaneously capture extended dynamic range and higher spectral resolution. We have also proposed a demosaicing algorithm that reduces aliasing artifacts. Our simulation results are based upon real multispectral images. They show that the combination of the GAP mosaic with submicrometer pixels and our simple demosaicing algorithm works well.

Our demosaicing algorithm, however, has the limitation that aliasing reduction cannot be applied when either a primary or a secondary channel is saturated. We plan to investigate this issue in our future work. We are also exploring the fabrication of a GAP sensor. An interesting challenge is to find the pigments needed to realize a CFA with the optimal spectral responses we have estimated.

## REFERENCES

- [1] R. Lyon and P. Hubel, "Eyeing the camera: Into the next century," in *Proc. IS&T/TSID 10th Color Imaging Conf.*, 2002, pp. 349–355.
- [2] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proc. IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition*, 2000, vol. 1, p. 1472.
- [3] S. G. Narasimhan and S. K. Nayar, "Enhancing resolution along multiple imaging dimensions using assorted pixels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 518–530, Apr. 2005.

- [4] J. F. H. JR and J. T. Compton, "Processing Color and Panchromatic Pixels," U.S. Patent: 2007/0024879 A1, Feb. 2007.
- [5] R. Shogenji, Y. Kitamura, K. Yamada, S. Miyatake, and J. Tanida, "Multispectral imaging using compact compound optics," *Opt. Exp.*, pp. 1643–1655, 2004.
- [6] G. A. Baone and H. Qi, "Demosaicking methods for multispectral cameras using mosaic focal plane array technology," in *Proc. SPIE*, Jan. 2006, vol. 6062.
- [7] K. Fife, A. E. Gamal, and H.-S. P. Wong, "A 3 Mpixel multi-aperture image sensor with 0.7  $\mu\text{m}$  pixels in 0.11  $\mu\text{m}$  CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf. Digest of Technical Papers*, Feb. 2008, pp. 48–594.
- [8] D. Alleysson, S. Süsstrunk, and J. Héroult, "Linear demosaicking inspired by the human visual system," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 439–449, Apr. 2005.
- [9] E. Dubois, "Filter design for adaptive frequency-domain bayer demosaicking," in *Proc. IEEE Int. Conf. Image Processing*, 2006, pp. 2705–2708.
- [10] K. Hirakawa and P. Wolfe, "Spatio-spectral color filter array design for optimal image recovery," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1876–1890, Oct. 2008.
- [11] M. J. Vrhel, R. Gershon, and L. S. Iwan, "Measurement and analysis of object reflectance spectra," *Color Res. Appl.*, vol. 19, 1994.
- [12] J. Parkkinen, J. Hallikainen, and T. Jaaskelainen, "Characteristic spectra of munsell colors," *J. Opt. Soc. Amer.*, vol. 6, Feb. 1989.
- [13] G. Sharma and H. Trussell, "Figures of merit for color scanners," *IEEE Trans. Image Process.*, vol. 6, no. 7, Jul. 1997.
- [14] S. Quan, N. Ohta, R. Berns, and X. Jiang, "Unified measure of goodness and optimal design of spectral sensitivity functions," *J. Imaging Sci. Technol.*, 2002.
- [15] J. I. Park, M. H. Lee, M. Crossberg, and S. K. Nayar, "Multispectral imaging using multiplexed illumination," in *Proc. IEEE Int. Conf. Computer Vision*, Oct. 2007, pp. 1–8.
- [16] G. C. Holst, *CCD Arrays, Cameras and Displays*, 2nd ed. Winter Park, FL: JCD Publishing, 1998.
- [17] [Online]. Available: <http://www.asahi-spectra.com/>
- [18] [Online]. Available: <http://www.anywhere.com/gward/hdrenc/pages/originals.html>
- [19] [Online]. Available: [http://www.cis.rit.edu/mcsl/icam/hdr/rit\\_hdr/](http://www.cis.rit.edu/mcsl/icam/hdr/rit_hdr/)
- [20] B. K. Gunturk, J. Glotzbach, Y. Altunbasak, R. W. Schafer, and R. M. Mersereau, "Demosaicking: Color filter array interpolation," *IEEE Signal Process. Mag.*, vol. 22, no. 1, pp. 44–54, Jan. 2005.
- [21] K. Hirakawa and T. Parks, "Adaptive homogeneity-directed demosaicking algorithm," in *Proc. IEEE Int. Conf. Image Processing*, Sep. 2003, vol. 3, pp. 669–672.
- [22] B. Gunturk, Y. Altunbasak, and R. Mersereau, "Color plane interpolation using alternating projections," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 997–1013, Sep. 2002.



**Fumihito Yasuma** received the B.E. degree in information and system engineering from the Chuo University, Japan in 2004, and M.E. degree in information science and systems engineering from the Ritsumeikan University, Japan in 2006. He has been working for Sony Corporation since 2006.

He studied as a visiting scholar with Prof. S. Nayar in Columbia University from 2007 to 2009.



**Tomoo Mitsunaga** received the B.E. and M.E. degree in biophysical engineering from Osaka University, Japan, in 1989 and 1991, respectively.

He has been working for Sony Corporation, Tokyo, Japan, since 1991. He studied as a visiting scholar with Prof. Shree Nayar at Columbia University from 1997 to 1999. His interests include computer vision, digital image processing, and computational cameras.



**Daisuke Iso** received the B.E, M.E, and Ph.D. degrees in information and computer science from Keio University, Tokyo, Japan, in 2001, 2003, and 2006, respectively.

He has been working for Sony Corporation, Tokyo, Japan, since 2003. His research interests include computer vision, image processing, and computational photography.



**Shree K. Nayar** (S'86–M'90) received the Ph.D. degree in electrical and computer engineering from the Robotics Institute at Carnegie Mellon University, Pittsburgh, PA, in 1990.

He is currently the T. C. Chang Professor of Computer Science at Columbia University, New York. He co-directs the Columbia Vision and Graphics Center. He also heads the Columbia Computer Vision Laboratory (CAVE), which is dedicated to the development of advanced computer vision systems. His research is focused on three areas; the creation of novel

cameras, the design of physics based models for vision, and the development of algorithms for scene understanding. His work is motivated by applications in the fields of digital imaging, computer graphics, and robotics.

Dr. Nayar has received best paper awards at ICCV 1990, ICPR 1994, CVPR 1994, ICCV 1995, CVPR 2000, and CVPR 2004. He is the recipient of the David Marr Prize (1990 and 1995), the David and Lucile Packard Fellowship (1992), the National Young Investigator Award (1993), the NTT Distinguished Scientific Achievement Award (1994), the Keck Foundation Award for Excellence in Teaching (1995), the Columbia Great Teacher Award (2006), and Carnegie Mellon University's Alumni Achievement Award (2009). In February 2008, he was elected to the National Academy of Engineering.